Causal Markov mesh hierarchical modeling for the contextual classification of multiresolution satellite images

Alessandro Montaldo, Luca Fronda, Ihsen Hedhli, Gabriele Moser, Sebastiano B. Serpico, Josiane Zerubia

CAUSAL MARKOV MESH HIERARCHICAL MODELING FOR THE CONTEXTUAL CLASSIFICATION OF MULTIRESOLUTION SATELLITE IMAGES

A. Montaldo[1]*, L. Fronda*[1]*, I. Hedhli*[2]*, G. Moser*[1]*, S. B. Serpico*[1]*, J. Zerubia*[3]

[1] University of Genoa, Italy, gabriele.moser@unige.it. [2] Université Laval, Quebec, Canada
[3] Inria, Sophia-Antipolis Méditerranée center, and Université Cote d'Azur, France

ABSTRACT

In this paper, we address the problem of the joint classification of multiple images acquired on the same scene at different spatial resolutions. From an application viewpoint, this problem is of importance in several contexts, including, most remarkably, satellite and aerial imagery. From a methodological perspective, we use a probabilistic graphical approach and adopt a hierarchical Markov mesh framework that we have recently developed and models the spatial-contextual classification of multiresolution and possibly multisensor images. Here, we focus on the methodological properties of this framework. First, we prove the causality of the model, a highly desirable property with respect to the computational cost of the inference. Then, we prove the expression of the marginal posterior mode criterion for this model and discuss the related assumptions. Experimental results with multispectral and panchromatic satellite images are also presented.

***Index Terms*—** Multiresolution images, causality, hierarchical Markov random field, Markov mesh random field, semantic image segmentation.

1. INTRODUCTION

The joint availability of images acquired on the same scene at different spatial resolutions is a common scenario in many applications. Examples with optical data include panchromatic, multispectral, and hyperspectral images taken by sensors (or ensembles of sensors) onboard satellites, aircrafts, or drones [1, 2], as well as recent dedicated multiresolution cameras [3, 4]. In the case of radar imagery, different synthetic aperture modalities (e.g., stripmap, spotlight, ScanSAR) allow collecting data with various tradeoffs between resolution and coverage [5]. From an image analysis perspective, the challenge is to develop processing methods that benefit from all available resolutions and model the tradeoff between the synoptic view of the coarser ones and the spatial detail of the finer ones.

The focus of this paper is on the (dense) supervised classification (or semantic segmentation [6, 7]) of multiresolution images. It is a challenging and scarcely investigated task, for which common approaches mostly make use of resampling procedures, which are computationally cheap but may generally yield artifacts [1]. The approach used here is based on hierarchical latent Markov modeling [8]. We focus on the framework that we recently developed in [9] and we investigate its methodological properties in terms of causality of the stochastic model and analytical formulation of the inference.

We recall that, in many image processing applications, Markov random fields (MRFs) on either planar or multilayer graphs have been popular for long as powerful stochastic models for spatial and possibly multimodal information [8]. However, a shortcoming of a generic MRF is that it is generally noncausal. This is no restriction from a modeling perspective but leads to iterative (e.g., stochastic or graph-theoretic) inference algorithms, which may be time-consuming, especially as compared to the 1D case of Markov chains. Among the broad family of MRFs, two sub-classes of models for which causality can be formalized include Markov mesh random fields (MMRFs) on planar lattices [10] and hierarchical MRFs on quadtrees [11]. In the former case, a case-specific notion of neighborhood is formulated on a planar pixel grid so that causality is proven to hold [10, 12, 13, 14, 15]. The latter are associated with a quadtree topology and are formalized in terms of a Markovianity property across the layers of the tree (i.e., along the spatial scale) [11, 16]. On one hand, for both families of models, causality makes it possible to formulate efficient inference algorithms. On the other hand, the two families exhibit complementary properties: an MMRF describes spatial interactions among the pixels but is intrinsically a single-resolution model, while a hierarchical MRF on a quadtree captures multiresolution relations across the layers of the tree but does not model the spatial context within each layer.

In [9], we have developed a hierarchical MMRF that integrates both modeling approaches in a unique framework. In this model, Markovianity is postulated both across the scales of a quadtree and with respect to the neighborhood system of a mesh associated with each layer of the tree. This joint strategy benefits from the spatial information within each layer and inherently supports multiresolution fusion. The model has been combined with both Gaussian mixtures in the application to multiresolution optical data [9] and with decision tree ensembles in the application to multiresolution and mul-

tisensor (optical and radar) imagery [17]. In this framework, the present paper addresses the methodological properties of this combined hierarchical Markov mesh. The main contributions are twofold. First, we prove the causality of the model. While causality is a well-known result for a hierarchical MRF and for a planar MMRF, it is not *a priori* guaranteed for their combination. We prove here that it holds for our integrated framework as well. Secondly, we prove the analytical formulation of the marginal posterior mode (MPM) criterion for this model and we discuss the related conditional independence assumptions. MPM inference is especially advantageous for classification and segmentation methods associated with multiresolution models [11]. Its formulation for the hierarchical MMRF explicitly relies on the causality of the model. Examples of experimental results obtained within this framework, in addition to those in [9, 17], are shown in a case study associated with high resolution satellite multiresolution data of an urban and agricultural area and with the application to land cover mapping for flood risk management [18].

2. METHODOLOGY

2.1. Causal hierarchical Markov mesh image model

An image classification problem can be regarded as the process that estimates the latent information x (class label) from the observation y attached to one or more nodes (i.e., pixels) s. In statistical approaches, x and y are usually viewed as occurrences of random vectors $\mathcal{X}$ and $\mathcal{Y}$. However, inference of $\mathcal{X} = x$ given $\mathcal{Y} = y$ is computationally demanding and in most cases, iterative algorithms are required. For that, probabilistic causal image models have been studied since the early 1990's through hierarchical MRFs on quadtrees. These models rely on a causality concept captured by the factorization of the prior distribution (i.e., the distribution of $\mathcal{X}$) in terms of causal transition probabilities. Let $\{S^0, S^1, \ldots, S^L\}$ be a set of pixel grids arranged as a quadtree. Hence, the width and height of $S^{\ell-1}$ are half those of S^ℓ, and each site $s \in S^\ell$ has a parent site $s^- \in S^{\ell-1}$ and four children sites in $S^{\ell+1}$ (collected in a set $s^+ \subset S^{\ell+1}; \ell = 1, 2, \ldots, L-1$). A hierarchy on the tree $S = \bigcup_{\ell=0}^{L} S^\ell$ from the root to the leaves is determined [11]. If a discrete class label x_s is associated with each $s \in S$, then $\mathcal{X} = \{x_s\}_{s\in S}$ is a hierarchical MRF if [11, 16]:

$$P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}, \mathcal{X}^{\ell-2}, \ldots, \mathcal{X}^0) = P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}), \qquad (1)$$

where $\mathcal{X}^\ell = \{x_s\}_{s\in S^\ell}$ ($\ell = 1, 2, \ldots, L$), i.e., if Markovianity holds across the scales. In this hierarchical model, these transition probabilities also factorize so that [11] (see Fig. 1(a)):

$$P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}) = \prod_{s\in S^\ell} P(x_s|x_{s^-}). \qquad (2)$$

This condition removes the contextual dependency within $\mathcal{X}^\ell$.

For the observation model $P(\mathcal{Y}|\mathcal{X})$, where $\mathcal{Y} = \{y_s\}_{s\in S}$ is the random field of the observations associated with all sites of the tree, a standard site-wise factorization is assumed:

$$P(\mathcal{Y}|\mathcal{X}) = \prod_{s\in S} P(y_s|x_s) = \prod_{\ell=0}^{L} \prod_{s\in S^\ell} P(y_s|x_s). \qquad (3)$$

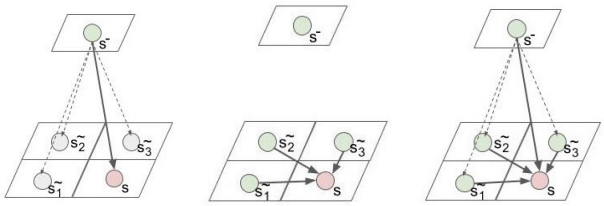

Fig. 1. (a) Structural causal models used in (a) a hierarchical MRF, (b) an MMRF, and (c) the hierarchical MMRF.

In the considered framework, the quadtree as a structural causal model is extended to a multiscale model in which spatial information is incorporated while keeping the causality of the hierarchical model. For that, let us mention another important class of causal MRFs, i.e., the MMRFs on a rectangular lattice R. An MMRF is build upon an order relation $\prec$ on R, so that, intuitively, the "past" of each site $s \in R$ (i.e., the sites $r \in R$ such that $r \prec s$) is well-defined. A neighborhood relation is assumed in R consistently with this order relation, and $r \precsim s$ indicates that r is a causal neighbor of s. $\mathcal{X} = \{x_s\}_{s\in R}$ is an MMRF if, for all $s \in R$, [10, 12, 13]:

$$P(x_s|x_r, r \prec s) = P(x_s|x_r, r \precsim s), \qquad (4)$$

i.e., if Markovianity holds when restricted to the past of each site. A common choice is that the past of s is the set of all pixels traversed by a raster scan before reaching s and that the neighbors of s are the three adjacent pixels located in its previous row and column (see Fig. 1(b)). With this choice, the following factorization holds (up to appropriate definition of the behavior at the image borders) [13]:

$$P(\mathcal{X}) = \prod_{s\in R} P(x_s|x_r, r \precsim s). \qquad (5)$$

In the considered hierarchical MMRF framework, the MMRF dependencies in (5) are taken into account so that the transition probabilities of the Markov chain in (1) factorize in such a way that the components of $\mathcal{X}^\ell$ express both the past (spatial) and the parent-child dependencies. More precisely, in the hierarchical MMRF framework on the quadtree S, we assume that: (i) $\mathcal{X}$ satisfies the hierarchical Markovianity in (1); (ii) $\mathcal{X}^0$ is an MMRF on the root lattice S^0; (iii) the following proportionality holds ($\ell = 1, 2, \ldots, L$):

$$P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}) \propto \prod_{s\in S^\ell} P(x_s|x_r, r \precsim s) P(x_s|x_{s^-}); \qquad (6)$$

and (iv) $\mathcal{Y}$ satisfies the conditional independence in (3).

Here, we prove the causality of this combined framework. According to (1), we obtain:

$$P(\mathcal{X},\mathcal{Y}) = P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}) = P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L,\mathcal{X}^{L-1},\dots,\mathcal{X}^0)$$
$$= P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L|\mathcal{X}^{L-1})\dots P(\mathcal{X}^1|\mathcal{X}^0)P(\mathcal{X}^0). \quad (7)$$

Plugging the factorizations (3) for $P(\mathcal{Y}|\mathcal{X})$, (5) for $P(\mathcal{X}^0)$, and (6) for $P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1})$ ($\ell = 1,2,\dots,L$) into (7) yields:

$$P(\mathcal{X},\mathcal{Y}) \propto \prod_{\ell=1}^{L}\prod_{s\in S^\ell} P(x_s|x_r, r \lesssim s)P(x_s|x_{s^-})P(y_s|x_s)\cdot$$
$$\cdot \prod_{s\in S^0} P(x_s|x_r, r \lesssim s)P(y_s|x_s). \quad (8)$$

Therefore, $P(\mathcal{X},\mathcal{Y})$ is entirely defined by the parent-child transition probabilities $P(x_s|x_{s^-})$, the causal sibling transition probabilities $P(x_s|x_r, r \lesssim s)$ and the data conditional likelihoods $P(y_s|x_s)$. This factorization implies that $(\mathcal{X},\mathcal{Y})$ is a Markov random field with respect to the causal structure for the prior distribution defined in Fig.1(c), which determines the causality of the hierarchical MMRF framework.

2.2. Inference algorithm and MPM criterion

The hierarchical MMRF is causal both spatially and across scales, which allows an efficient recursive algorithm to be formulated for the MPM criterion. MPM assigns each $s \in S$ the class label x_s that maximizes $P(x_s|\mathcal{Y})$ [8]. Here, we prove that, under suitable conditional independence assumptions (see below), the following relations hold ($s \in S$):

$$P(x_s) = \sum_{x_{s^-}} P(x_s|x_{s^-})P(x_{s^-}), \quad (9)$$

$$P(x_s|y_s^d) \propto P(x_s|y_s) \prod_{t\in s^+}\sum_{x_t} \frac{P(x_t|y_t^d)P(x_t|x_s)}{P(x_t)}, \quad (10)$$

$$P(x_s|x_s^c, y_s^d) \propto \frac{P(x_s|y_s^d)P(x_s|x_{s^-})P(x_{s^-})}{P(x_s)^{n_s}}\cdot$$
$$\cdot \prod_{r\lesssim s} P(x_s|x_r)P(x_r), \quad (11)$$

$$P(x_s|\mathcal{Y}) = \sum_{x_s^c} P(x_s|x_s^c, y_s^d)P(x_{s^-}|\mathcal{Y}) \prod_{r\lesssim s} P(x_r|\mathcal{Y}), \quad (12)$$

where y_s^d collects the observations of all descendants of s in the tree (including s itself), x_s^c collects the labels of all sites connected to s (i.e., x_{s^-} and $\{x_r\}_{r\lesssim s}$), and n_s is the number of such sites. Accordingly, MPM inference in the considered framework is accomplished through three recursive steps. First, (9) is used to calculate $P(x_s)$ on all sites through a top-down pass from the root to the leaves. Then, (10) and (11) are used to compute $P(x_s|x_s^c, y_s^d)$ through a bottom-up pass from the leaves to the root. Finally, (12) is used to derive $P(x_s|\mathcal{Y})$ through a second top-down pass.

Details on the initialization of these recursions and on the parametric modeling of the transition probabilities $P(x_s|x_{s^-})$ and $P(x_s|x_r), r \lesssim s$, can be found in [9]. Here, we focus on deriving (9)-(12) and on the related assumptions.

Specifically, (9) is a straightforward application of the total probability theorem, and the proof of (10) in the case of the hierarchical MMRF is identical to that reported in [11] for a hierarchical MRF. (11) and (12) hold under the following conditional independence assumptions:

$$\text{A1}: \quad P(x_s|x_s^c,\mathcal{Y}) = P(x_s|x_s^c, y_s^d) \quad (13)$$
$$\text{A2}: \quad P(x_s^c|\mathcal{Y}) = P(x_{s^-}|\mathcal{Y}) \prod_{r\lesssim s} P(x_r|\mathcal{Y})$$
$$\text{A3}: \quad P(x_s^c|x_s, y_s^d) = P(x_s^c|x_s) = P(x_{s^-}|x_s) \prod_{r\lesssim s} P(x_r|x_s)$$

Indeed, the total probability theorem implies that:

$$P(x_s|\mathcal{Y}) = \sum_{x_s^c} P(x_s|x_s^c,\mathcal{Y})P(x_s^c|\mathcal{Y}), \quad (14)$$

and (12) follows from plugging A1 and A2 into this equation. Moreover, thanks to Bayes theorem:

$$P(x_s|x_s^c, y_s^d) \propto P(x_s^c|x_s, y_s^d)P(x_s|y_s^d), \quad (15)$$

where the proportionality constant does not depend on x_s. Plugging A3 into this relation implies:

$$P(x_s|x_s^c, y_s^d) \propto P(x_s|y_s^d)P(x_{s^-}|x_s) \prod_{r\lesssim s} P(x_r|x_s), \quad (16)$$

from which (11) follows due to Bayes theorem.

Assumption A1 means that the label of s, given the parent and sibling labels, only depends on the observations of the descendants of s and not on those of the other sites. A2 implies that, given the observation field, the parent and the sibling labels of s are conditionally independent. A3 means that the parent and sibling labels of s, when conditioned to the label of s, are independent of the observations of the descendants of s and mutually independent. These statements are similar to the conditional independence conditions that are usually accepted for analytical convenience in the case of hierarchical [11] or planar MRFs [8].

In particular, the symmetric mesh in [14] is used to prevent anisotropic artifacts. In the case of a planar lattice, it is based on an appropriate pixel visiting scheme, which ensures corner independence [14]. In the hierarchical MMRF, this visiting scheme is applied within each layer of the quadtree when it is reached by the aforementioned recursive steps.

The observations enter the recursions through the pixelwise posteriors $P(x_s|y_s)$ in (10). Decision tree ensembles, including random forest [19], rotation forest [20], Extra-Trees [21], and gradient boosted regression trees (GBRT) [22], are used to estimate these pixelwise posteriors based on the training samples of the classes. As discussed in [17, 23], their nonparametric formulation allows for both single- and multisensor data to be integrated into the hierarchical MMRF.

Table 1. Overall accuracy and class-by-class accuracies on the test set for the hierarchical MMRF framework, applied in conjunction with random forest (RanFor), rotation forest (RotFor), ExtraTrees, and GBRT, and for the previous method in [24].

resolution 1 m	urban %	low vegetation 1 %	low vegetation 2 %	bare soil %	tall vegetation %	overall %	Cohen's kappa coeff
Proposed method, RanFor	99.01	81.40	79.77	94.22	74.26	85.15	0.7874
Proposed method, RotFor	98.07	73.34	43.23	91.65	71.12	72.44	0.6127
Proposed method, ExtraTrees	98.94	84.50	81.76	95.08	74.91	87.18	0.8149
Proposed method, GBRT	**98.94**	**84.55**	**81.76**	**95.19**	**74.50**	**87.21**	**0.8152**
Method in [24]	97.75	62.33	82.14	82.25	98.62	76.02	0.6744
resolution 2 m	urban %	low vegetation 1 %	low vegetation 2 %	bare soil %	tall vegetation %	overall %	Cohen's kappa coeff
Proposed method, RanFor	98.98	84.93	79.88	94.60	70.46	86.67	0.8108
Proposed method, RotFor	97.36	78.49	39.89	81.86	60.06	73.70	0.6258
Proposed method, ExtraTrees	98.97	88.26	82.13	96.70	72.82	88.90	0.8421
Proposed method, GBRT	**98.97**	**88.25**	**82.13**	**96.70**	**72.82**	**88.90**	**0.8421**
resolution 4 m	urban %	low vegetation 1 %	low vegetation 2 %	bare soil %	tall vegetation %	overall %	Cohen's kappa coeff
Proposed method, RanFor	99.13	93.04	80.28	91.12	50.00	90.11	0.8660
Proposed method, RotFor	97.19	78.26	50.82	82.84	70.83	75.54	0.6590
Proposed method, ExtraTrees	99.10	95.49	82.67	94.08	52.78	91.94	0.8935
Proposed method, GBRT	**99.10**	**95.49**	**82.67**	**94.08**	**52.78**	**91.94**	**0.8935**

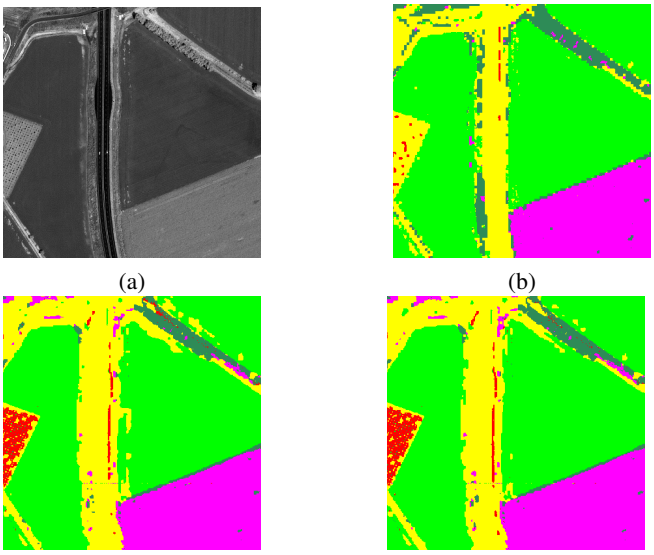

Fig. 2. Details of (a) the IKONOS panchromatic image at 1-m resolution and of the maps obtained at the same resolution by (b) the method in [24] and by the proposed framework, when applied with (c) random forest and (d) GBRT.

3. EXPERIMENTAL VALIDATION

Experimental results are shown in the application to an optical satellite data set acquired by the IKONOS mission in 2004 over the area of Alessandria, Italy. It consists of a single-channel panchromatic image at 1-m resolution (1260×1400 pixels) and a 4-channel (RGB and near infrared) multispectral image at 4-m resolution (315×350 pixels). This data set and its classification were framed within a case study of flood vulnerability assessment [18]. Based on the aforementioned resolutions, a quadtree with three layers was used, including the panchromatic and multispectral images in the leaf and root layers, respectively. The intermediate layer was filled in by pansharpening through the Gram-Schmidt algorithm [25] and resampling the pansharpened image onto a 2-m pixel grid. Non-overlapping training and test sets were manually annotated by a specialist. The multiresolution optical image classification method in [24], based on noncausal MRFs, Gaussian processes, and linear mixtures, was used for comparison purposes. The accuracies obtained on the test set are shown in Table 1. Details of the classification maps are in Fig. 2.

4. DISCUSSION AND CONCLUSION

The proposed approach obtained quite high accuracies on the test set in all layers of the quadtree, despite the strong spectral overlapping between several classes, which corresponded to vegetation covers observed through only a few spectral channels. The resulting map exhibited remarkable visual regularity (Fig. 2), thus suggesting the effectiveness of the adopted spatial-contextual mesh. The results in Table 1 also indicate the flexibility of our approach in incorporating pixelwise predictions from different tree ensembles. In particular, the highest accuracies were achieved with GBRT and ExtraTrees. The technique in [24] also obtained a rather satisfactory discrimination of most classes, although with significantly lower accuracies than the proposed framework. Moreover, this previous technique maps only at the finest scale while the proposed framework simultaneously classifies at all three scales.

From a data fusion perspective, these results confirm the effectiveness of addressing multiresolution classification through a causal hierarchical Markov mesh framework, even as compared to a previous approach based on noncausal Markov modeling and linear mixtures. For the hierarchical MMRF, both causality and the formulation of MPM have been proven analytically. The experimental results shown here with optical imagery and in [9, 17] with also multisensor data confirm its effectiveness in different scenarios of multiresolution image classification with satellite imagery.

5. REFERENCES

[1] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*, CRC Press, 2008.

[2] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2882–2895, 2016.

[3] G. Y. Belay, H. Ottevaere, Y. Meuret, M. Vervaeke, J. Van Erps, and H. Thienpont, "Demonstration of a multichannel, multiresolution imaging system," *Applied Optics*, vol. 52, no. 24, pp. 6081–6089, 2013.

[4] X. Lu and X. Li, "Multiresolution imaging," *IEEE Transactions on Cybernetics*, vol. 44, no. 1, pp. 149–160, 2014.

[5] D. Massonnet and J.-C. Souyris, *Imaging with synthetic aperture radar*, EPFL Press, 2008.

[6] A. Arnab, S. Zheng, S. Jayasumana, B. Romera-Paredes, M. Larsson, A. Kirillov, B. Savchynskyy, C. Rother, F. Kahl, and P. H. S. Torr, "Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 37–52, 2018.

[7] X. Wang, "Deep learning in object recognition, detection, and segmentation," *Foundations and Trends in Signal Processing*, vol. 8, no. 4, pp. 217–382, 2016.

[8] S.Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer, 2009.

[9] I. Hedhli, G. Moser, S. B. Serpico, and J. Zerubia, "Multi-resolution classification of urban areas using hierarchical symmetric Markov mesh models," in *Proc. of JURSE 2017*, 2017.

[10] P. A. Devijver, "Hidden Markov mesh random field models in image analysis," *J. Applied Stat.*, vol. 20, no. 5-6, pp. 187–227, 1993.

[11] J. . Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 390–404, 2000.

[12] A. P. Dunmur and D. M. Titterington, "Computational Bayesian analysis of hidden Markov mesh models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1296–1300, 1997.

[13] K. Abend, T. J. Harley, and L. N. Kanal, "Classification of binary random patterns," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 538–544, 1965.

[14] S. Yousefi, N. Kehtarnavaz, and Y. Cao, "Computationally tractable stochastic image modeling based on symmetric Markov mesh random fields," *IEEE Trans. Image Process.*, vol. 22, no. 6, 2013.

[15] Q. R. Razlighi, N. Kehtarnavaz, and A. Nosratinia, "Computation of image spatial entropy using quadrilateral Markov random field," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2629–2639, 2009.

[16] A. S. Willsky, "Multiresolution Markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, 2002.

[17] A. Montaldo, L. Fronda, I. Hedhli, G. Moser, J. Zerubia, and S. B. Serpico, "Joint classification of multiresolution and multisensor data using a multiscale markov mesh model," in *IEEE IGARSS 2019 (in print)*, 2019.

[18] S. B. Serpico, S. Dellepiane, G. Boni, G. Moser, E. Angiati, and R. Rudari, "Information extraction from remote sensing images for flood monitoring and damage evaluation," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2946–2970, 2012.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, 2006.

[21] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[22] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[23] A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*, Springer, 2013.

[24] G. Moser, A. De Giorgi, and S. B. Serpico, "Multiresolution supervised classification of panchromatic and multispectral images by Markov random fields and graph cuts," *IEEE Trans. Geosci. Rem. Sens.*, vol. 54, no. 9, pp. 5054–5070, 2016.

[25] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 446–453, 2011.